



Digitale Annotationen: 'Best Practices' und Potentiale (R 6.2.1) Teil II

Version 29. Februar 2016

Cluster 6.2 Fachwissenschaftliche Annotationen – Methoden

Verantwortliche Partner BBAW, STI, TUD

DARIAH-DE Aufbau von Forschungsinfrastrukturen für die e-Humanities

Dieses Forschungs- und Entwicklungsprojekt wird / wurde mit Mitteln des Bundesministeriums für Bildung und Forschung (BMBF), Förderkennzeichen 01UG1110A bis N, gefördert und vom Projektträger im Deutschen Zentrum für Luft- und Raumfahrt (PT-DLR) betreut.

GEFÖRDERT VOM



**Bundesministerium
für Bildung
und Forschung**

Projekt: DARIAH-DE: Aufbau von Forschungsinfrastrukturen für die e-Humanities

BMBF Förderkennzeichen: 01UG1110A bis N

Laufzeit: März 2011 bis Februar 2016

Dokumentstatus: final

Verfügbarkeit: öffentlich

Autoren:

Rainer Becker (TUD), Michael Bender (TUD), Luise Borek (TUD), Canan Hastik (TUD), Thomas Kollatz (STI), Beata Mache (STI), Harald Lordick (STI), Ruth Reiche (TUD)

Revisionsverlauf:

Datum	Autor	Kommentare
29. Februar 2016	Becker, Rainer (TUD)	
29. Februar 2016	Bender, Michael (TUD)	
29. Februar 2016	Borek, Luise (TUD)	
29. Februar 2016	Hastik, Canan (TUD)	
29. Februar 2016	Kollatz, Thomas (STI)	
29. Februar 2016	Mache, Beata (STI)	
29. Februar 2016	Lordick, Harald (STI)	
29. Februar 2016	Reiche, Ruth (TUD)	

Inhaltsverzeichnis:

Einleitung	5
Nutzungsszenarien digitaler Annotation	6
DARIAH-DE Annotation Sandbox	6
Annotieren im Web	8
Verfahren der Web-Annotation	9
Anwendungsszenarien.....	10
Ausblick	13
Digitale Annotationen und das Semantic Web.....	14
Text	16
Projekte der Infrastruktur Webanno	19
Bild.....	21
Video	23
Objekte.....	26
Noten	26
Literatur	27
Allgemein.....	27
Webannotation	27
Semantic Web.....	27
Text	27
Bild.....	28
Video	28

Einleitung

Während der erste Teil des Reports eine theoretisierende Perspektive einnimmt und damit aktuelle Diskussionen aus der europäischen Forschungsdebatte (z.B. aus dem Kontext der DARIAH-EU Working Group Digital Annotations) aufgreift, widmet sich der hier vorliegende zweite Teil praktischen Annotationsszenarien aus der deutschsprachigen Forschungslandschaft. Ausgehend von dem Einsatz konkreter Werkzeuge sowie fachspezifischer Fragestellungen werden exemplarisch Lösungen und Herausforderungen des digitalen Annotierens aufgezeigt und Überlegungen zu einer optimierten und nicht zuletzt nachhaltigen Verwendung resultierender Annotationen angestellt. Dabei wird auf Beispiele aus dem direkten Umfeld von DARIAH-DE ebenso zurückgegriffen wie auf allgemeinere Kontexte, die sich mit der Annotation verschiedenster Objekttypen befassen.

Dem Text gehen Vorarbeiten aus der zweiten Förderphase von DARIAH-DE voraus, z.B. eine Aufstellung verschiedener bestehender Annotationstools¹ und eine Spezifikation von Diensten, die sich für die Integration in die DARIAH-DE-Infrastruktur eignen². Er knüpft zudem an den Dialog mit der deutschsprachigen Forschungscommunity an, wie er z.B. in diversen Round Tables zum digitalen Annotieren sowie innerhalb des kürzlich bei der Gesellschaft für Informatik eingerichteten Arbeitskreises „Wissenschaftliche Annotationstechniken“ geführt wird.

Mit der DARIAH-DE Annotation Sandbox wird zunächst ein neuer Service vorgestellt, der sich als Resultat des o.g. Meilensteins präsentiert und Annotator.js mit dem DARIAH-DE Repository verknüpft. Ein anschließender Abschnitt zu Annotationen im Web liefert einen erweiterten (Über-)Blick auf den Bereich der Web-Annotation, der schließlich in eine Reflexion zu Semantic-Web-Strategien mündet. Ausgehend von Einblicken in die Annotationspraxis textueller Gegenstände widmen sich die folgenden Abschnitte weiteren Objekttypen wie Bild und Video. Die hier verwendete Literatur findet sich als Beitrag in die von DARIAH-DE gepflegte Doing Digital Humanities Bibliography verzeichnet, die einen gesonderten Bereich für das Annotationsthema enthält.

¹ https://docs.google.com/spreadsheets/d/1pScqOukgg3gLkL51fA_VUw9ubs_I08NwNJde-g3LEis/edit#gid=0

² <https://wiki.de.dariah.eu/x/Sg8CAg>

Nutzungsszenarien digitaler Annotation

DARIAH-DE Annotation Sandbox

Eine Hauptaufgabe des DARIAH-DE Clusters 6.2 bestand darin, einen generischen Annotationsdienst zur Integration in die DARIAH-DE Infrastruktur vorzuschlagen. Die Wahl fiel auf das Annotationstool Annotator.js.

Die Beweggründe, die dazu führten gerade dieses Werkzeug vorzuschlagen, sind ausführlich im DARIAH-DE Meilenstein Spezifikation von Diensten (M 6.2.1) erläutert.³

Kurz gesagt: Annotator.js ist ein "lightweight tool", gut dokumentiert, bereits verankert in der internationalen DH Community, verwendet von und gestützt durch viele und vielseitige darauf aufbauende Projekte und für ein breites Spektrum unterschiedlichster Annotationsszenarien insbesondere Text- und Bildannotation geeignet.

Nach Veröffentlichung des Meilensteins konnte die vorgeschlagene Integration von Annotator.js im Rahmen von zwei durch AP 6.2 initiierte DARIAH Programmiersprints – im Steinheim-Institut, Essen am 2. und 3. Februar 2015 und an der SUB Göttingen, am 18. und 19. Januar 2016 – realisiert werden. Die DARIAH-DE Annotation Sandbox⁴ kann in einer Betaversion seit Januar 2016 bereits genutzt werden.

Sie ermöglicht sowohl die Annotation von Texten und Bildern des TextGrid Repositoriums, als auch die Annotation beliebiger Webseiten, sowie die gezielte Einbettung des Annotationservices in eigene Webseiten.

Das TextGridRepositorium wurde mit einem Annotations-Login ausgestattet und ermöglicht nun sowohl Einzel- als auch Gruppenannotationen der im Repositorium vorhandenen Texteditionen.

Dieser Annotationsdienst ist sowohl für Textannotationen⁵ als auch für Annotation von Bilddigitalisaten⁶ eingerichtet. Letzteres wird durch ein Annotationsplugin der OpenSource Software DigiLib ermöglicht.

³ Veröffentlicht am 15.07.2014: "the annotation tool Annotator.js (...) to be adapted, further developed and integrated in the DARIAH-DE infrastructure"

⁴ <https://annotation.de.dariah.eu>

⁵ <https://annotation.de.dariah.eu/textgridrep/browse.html?id=textgrid:r0rg.0>

⁶ <https://annotation.de.dariah.eu/textgridrep/digilib/jquery/digilib.html?fn=textgrid:vqn8.0>

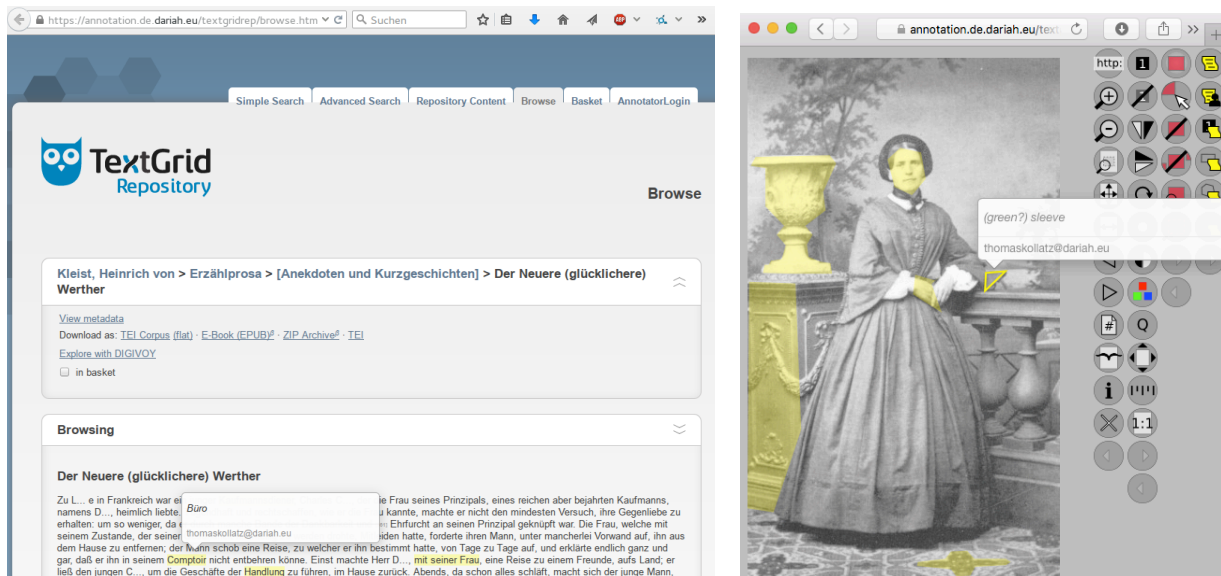


Abb.: Textannotation (links) – Bildannotation (rechts)

Darüber hinaus ist es möglich über die Webseite der DARIAH-DE Annotation Sandbox jede beliebige Website zu annotieren:

- DARIAH-DE Annotation Service
 - Annotate a Website

URL:

Abb.: Annotatieren einer Webseite

Als weitere Option kann der DARIAH-DE Annotationservice in eigene Webseiten eingebettet werden, indem die folgenden Zeilen in die Header-Sektion der Webseite eingebettet werden:

Embedding DARIAH-DE annotation service in your website

Add the following lines of code to the header of your website:

Start with jquery, if its not already included in in your website:

```
<script src="https://annotation.de.dariah.eu/annotatorjs/jquery-1.10.2.js"></script>
```

Afterwards add the following lines:

```
<link rel="stylesheet" type="text/css" href="https://annotation.de.dariah.eu/annotatorjs/annotator.min.css" />
<link rel="stylesheet" type="text/css" href="https://annotation.de.dariah.eu/annotatorjs/dariah-annotation.css" />
<script type="text/javascript" src="https://annotation.de.dariah.eu/annotatorjs/annotator-full.min.js"></script>
<script type="text/javascript" src="https://annotation.de.dariah.eu/annotatorjs/dariah-annotation.js"></script>
<script>
  DariahAnnotation.setupAnnotator();
  DariahAnnotation.addSidebar();
</script>
```

Abb.: Einbetten des DARIAH-DE Annotation Service in eine Webseite

Soll der Annotationsdienst nur eingeloggten Nutzern zur Verfügung stehen, dann ist vor die Zeile:

```
DariahAnnotation.setupAnnotator();
```

die folgende Anweisung zu schreiben:⁷

```
DariahAnnotation.anonymousViewing(false);
```

Die auf eine der oben beschriebenen Weisen angelegten Annotationen werden im DARIAH-DE Annotationsstorage gespeichert und können im AnnotationManager⁸, sofern es sich um öffentliche Annotationen handelt, gegebenenfalls inklusive der den Annotationen beigegebenen Tags [=Labels] angezeigt werden.

Mit der Integration des Annotator.js in DARIAH-DE können nun – wenn auch vorerst als Betadienst – nicht nur Forschungsdaten, sondern auch Annotationen dauerhaft in einer sicheren Forschungsinfrastruktur verwaltet und archiviert werden. Hier sind einige Weiterentwicklungen denkbar:

Tags und Labels ermöglichen die gezielte gegebenenfalls projekt- und disziplinübergreifende Auswertung fachwissenschaftlicher Annotationen (sofern das Rollen- und Rechtemanagement der Annotation dies zulassen)

Die Anbindung der Annotationen an Ontologien wie z.B. TaDiRAH könnte die Nachnutzung von Annotationen wesentlich erhöhen

Integration von Annotation in die Generic Search

Weitere Nutzungsszenarien eines Annotationservice, der Teil der digitalen Infrastruktur ist, sind im Meilenstein 6.2.1 beschrieben.

Annotieren im Web

Mit der längst selbstverständlichen digitalen Veröffentlichung von geisteswissenschaftlichen Quellen, Forschungsdaten und -ergebnissen ist die Notwendigkeit der Neu- und Weiterentwicklung von zeitgemäßen Arbeitstechniken verbunden. Auch das Annotieren als Kulturtechnik sollte in digitalen (Online-) Dokumenten uneingeschränkt möglich sein. Das Notieren von Fundstellen und Anmerkungen, das "Festhalten" des Entdeckten, Erkannten oder Entzifferten – essentielle Merkmale des wissenschaftlichen Arbeitens – wollen direkt in der digitalen Online-Quelle angemerkt werden können, also dort platziert, wo sie eigentlich auch hingehören: an das referenzierte Dokument. Denn so las-

⁷ Ein Beispiel für die Einbettung des Annotationsdienstes findet sich auf der Webseite der epigraphischen Datenbank epidat <http://steinheim-institut.de/cgi-bin/epidat?id=hha-3361>

⁸ <https://annotation.de.dariah.eu/AnnotationManager/>

sen sich beliebige Quellen durch bidirektionale und punktgenaue Verknüpfungen zuverlässig vernetzen. »Eine Online-Ressource sollte den Benutzer durch „backlinks“ zu den referenzierenden Stellen in Werken / Datenbanken verweisen können.«⁹ Das Thema Web-Annotation hat vielversprechendes Potenzial und ist in reger Diskussion und Entwicklung. Es lässt breite, disziplin- und anwendungsübergreifende Nutzungsszenarien schon jetzt erkennen und mehr noch erwarten.

Verfahren der Web-Annotation

Entsprechende Tools wie Pundit¹⁰, Annotator.js¹¹ oder Hypothes.is¹² zielen auf die Annotation von über das Web erreichbaren Ressourcen und zeichnen sich durch ein einfaches, intuitives Benutzerinterface aus. Die grundlegende Funktion ist jeweils ähnlich oder vergleichbar. Annotierbar werden Webseiten durch die Einbettung von Javascript-Code. Entweder durch direktes Einbetten durch den Anbieter der Webseite, zur Zeit aber vor allem durch sogenannte Bookmarklets, durch Browser-Plugins / -Extensions oder durch einen Proxy-Service¹³. Web-Annotationen sind öffentlich sichtbar, oder, bei vorausgesetztem Benutzer-Account, optional auch "privat". Die Annotationen werden auf einem zentralen Server in einem 'Store' oder 'Stream' zusammengeführt.¹⁴

Für die wissenschaftliche Anwendung ist die infrastrukturelle Unterstützung eines Web-Annotationssystems allerdings unverzichtbar.¹⁵ Das betrifft unter anderem ausdifferenzierte Eigenschaften der Authentifizierung und Autorisierung (u.a. privates versus öffentliches Teilen, Gruppenrechte), die nachhaltige Datenhaltung von Web-Annotationen als Mikro- oder Nanopublikationen¹⁶, die präzise, eindeutige und dauerhafte Referenzierung, soweit möglich die Rückbindung an Quellformate in Repositorien (etwa an XML-Knoten, IDs oder Datenbankfelder), die Einbeziehung von semantischen Strukturen und Normdaten sowie die Kommentierung multipler Web-Quellen in einer Annota-

⁹ Martin Raspe: Perspektiven der Forschung – PDF? Digitale Bildwissenschaft zwischen gestern und morgen. Vortrag in der Sektion »Bedrohte Besitzstände, verlorene Werte? Die Geisteswissenschaft von der Kunst und die neuen digitalen Verfahren«, DHd 2014.

¹⁰ <http://thepund.it/>

¹¹ <http://annotatorjs.org/>

¹² <https://hypothes.is/>

¹³ Wie etwa <https://via.hypothes.is/>

¹⁴ Die Technologie lässt sich prinzipiell auch in Intranets, sogenannten virtuellen Forschungs-umgebungen oder, mit Einschränkungen, auch nur auf lokalen 'Devices' nutzen.

¹⁵ Siehe das Meilenstein-Dokument: Annotation – Spezifikation von Diensten (M 6.2.1), 2014 <https://wiki.de.dariah.eu/pages/viewpage.action?pageId=33689418>

¹⁶ Vgl. dazu: Nanopublication Guidelines: Concept Web Alliance Working Draft 15 December 2013 <http://www.nanopub.org/2013/WD-guidelines-20131215/>

tion ("multiple targets"). Und auch die Vielfalt der benötigten Formate ist eine Herausforderung: Die für die Geisteswissenschaften relevanten Online-Quellen / Repositorien basieren auf höchst unterschiedlichen Technologien,¹⁷ liefern unterschiedlichste Formate. Ein in der Praxis tatsächlich nutzbares Werkzeug muss diese Vielfalt überspannen können – Medienbrüche gilt es zu vermeiden.

Anwendungsszenarien

Web-Annotationstechniken sind in ganz unterschiedlichen Phasen des wissenschaftlichen Arbeitens anwendbar. Sie können den Forschungsprozess unterstützen, Redaktionsprozesse begleiten, eine breite Fachdiskussion nach der Veröffentlichung der Ergebnisse ermöglichen. Potenzial wird insbesondere auch, über die Forschung hinaus, in ihrer Anwendung in der Lehre gesehen.¹⁸

Im geisteswissenschaftlichen Forschungsprozess, der insbesondere auch auf der Analyse und Kommentierung unterschiedlicher Quellen basiert, ist die Zusammenführung der Hinweise, Rechercheergebnisse und der eigenen Anmerkungen unverzichtbar. Die im Laufe der Forschung wachsende Materialsammlung, die Diversität an Notizen und Verweisen kann auf diese Weise in ein einheitliches System zusammengefügt und im Idealfall durch Anwendung von Normdaten, kontrollierten Vokabularen oder auch domänenspezifischen Auszeichnungen strukturiert werden.

Webannotationen sind heute aber auch geeignet, diese zunächst eher individuelle Perspektive zu überschreiten. Schon während der Forschung – also vor der Veröffentlichung – kann es sehr wünschenswert sein, Quellen kollaborativ zu erschließen, also Annotationen nicht "privat" zu halten, sondern zu "teilen". Das mag das kollaborative Zusammenwirken einer zunächst geschlossenen (projektbezogenen) Arbeitsgruppe betreffen, kann aber je nach Bedarf auch gleich öffentlich erfolgen. Öffentliche Annotationen einer Fachcommunity im letzteren Sinne wären auch ein Beispiel für sichtbar werdende Forschungsdaten, die üblicherweise bisher "in der Schublade" blieben.¹⁹ Die

¹⁷ So dass im Praxistest noch längst nicht jede Webseite mit Web-Annotationstools funktioniert: Je mehr clientseitiger Programmcode (Javascript) auf einer Webseite im Einsatz ist, desto eher scheint es zur Zeit Funktionsproblemen mit den verfügbaren Web-Annotationstools zu geben.

¹⁸ Zu "Classroom"-Szenarien siehe u.a. zahlreiche Hinweise bei <https://hypothes.is/education/>

¹⁹ Als heuristisch angelegtes Beispiel siehe: Harald Lordick: Notizen zum Blog als Web-Annotation <http://djgd.hypotheses.org/532>

konsequente Nutzung von "tags" erlaubt dabei das thematische Filtern der gespeicherten Annotationen. Sowohl die Plattform "Hypothes.is" als auch die DARIAH-DE-Annotatorkomponente bieten dieses Szenario.²⁰

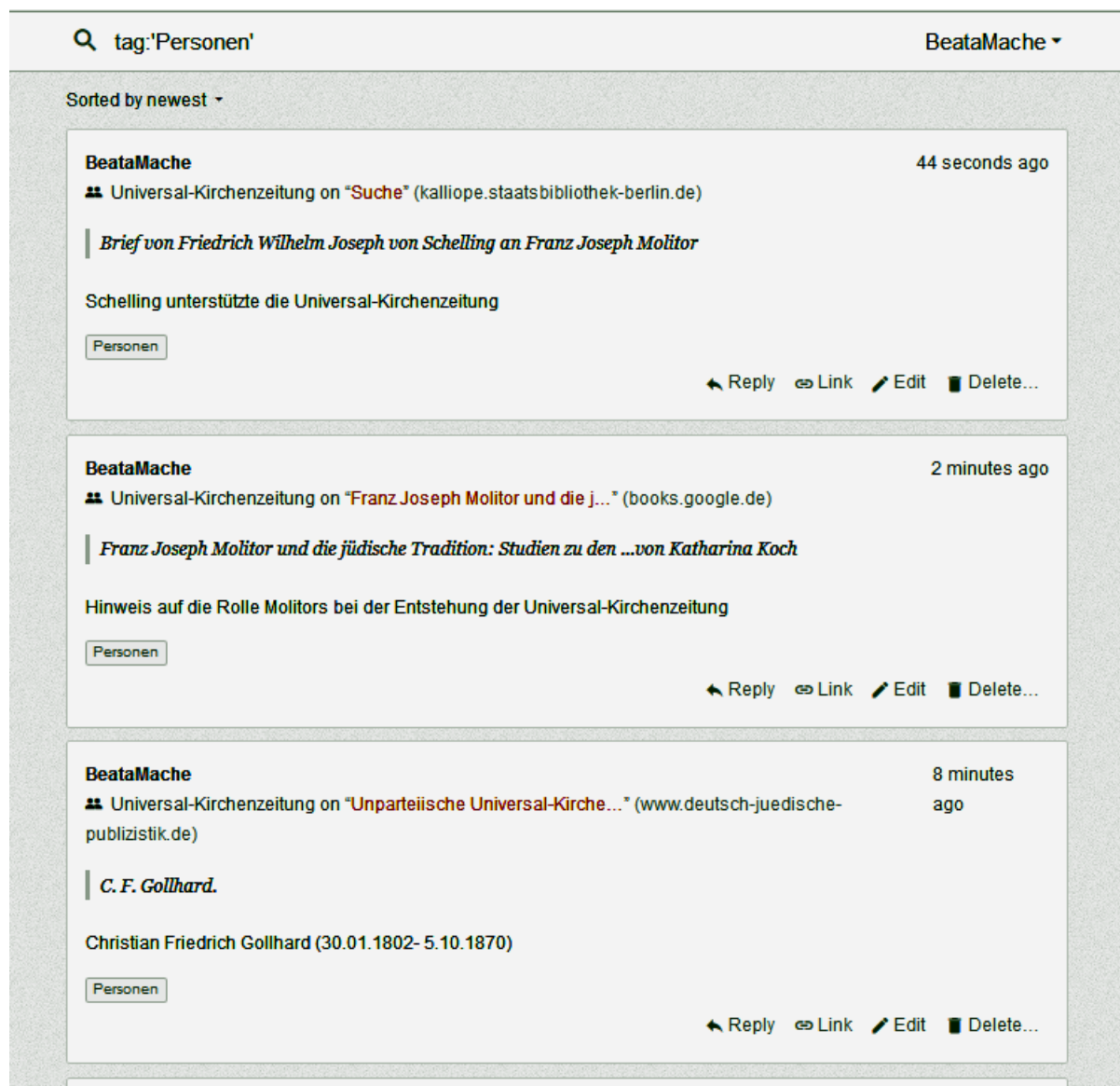


Abb.: Hypothes.is: Annotationen der Gruppe "Universal-Kirchenzeitung" im zusammengeführten Annotations-Stream, nach Label "Person" gefiltert

Die für die Geisteswissenschaften relevanten Quellenrepositorien bestehen nicht nur aus (maschinenlesbarem) Text (Volltext und Metadaten, als XML, HTML, PDF, Plain Text), sondern zunehmend auch aus digitalen Bild-, Audio- und Videoobjekten. Ein Annotationstool macht als generisches Werkzeug nur Sinn, wenn es möglichst breit solche

²⁰ Siehe nebenstehende Abbildung sowie den Abschnitt "Annotation Sandbox"

Formate unterstützt. Während "Text" mittlerweile weitgehend implementiert ist, besteht für die weiteren Formate noch erheblicher Entwicklungsbedarf.²¹

Abb.: Hypothes.is: Annotation von Metadaten im Onlinerepositorium

Besondere Beachtung verdienen die umfangreichen Sammlungen, deren Digitalisate (nur) aus gescannten Seitenabbildern (Images) bestehen. Die Unterstützung der Annotation solcher in Webseiten eingebetten Bilddigitalisate ist gerade wegen ihres quantitativ bedeutenden Anteils in den großen geisteswissenschaftlichen Repositorien (wie etwa Handschriften, frühe Drucke, Periodika) von großer Bedeutung.²²

Nicht zuletzt lässt sich gerade die manuelle Annotation dieser nicht maschinenlesbaren Seitenabbilder, wie sie für zahlreiche geisteswissenschaftlich relevante Repositorien typisch sind, als eine strukturierte Form der Erschließung von Quellen auffassen, an die wiederum computergestützte Suchalgorithmen, Auswertungs- ('Distant Reading') und Visualisierungsverfahren anknüpfen können.²³

²¹ Siehe auch die Abschnitte "Video" und "Bildannotation"

²² Vgl. Harald Lordick, Digitale Annotation im Spiegel praktischer Web-Annotation – Erschließung heterogener (Bild-) Quellen, DHd2015 <http://dx.doi.org/10.5281/zenodo.46422> . Als technologischer Rahmen für Web-Annotationsdienste, die sich für diese Repositorien eignen, steht das "International Image Interoperability Framework" (IIIF) bereit, auf dem etwa auch das von der DARIAH-DE Annotationskomponente (siehe "Annotation Sandbox") verwendete "digilib" basiert.

²³ Vgl. ders., Crossing Borders Through (Web) Annotations <http://dx.doi.org/10.5281/zenodo.32509>.

neuen Jahres 1837, bei *Franz Varrentrapp*:

Unparteiische Universal-Kirchenzeitung Für die Geistlichkeit und die gebildete Weltklasse des protestantischen, katholischen und israelitischen Deuts

In Verbindung mit

Archidiakonus Dr. *Constantin Ackermann* in Jena — Dr. *Adler* in Worms — Domfessor Dr. *Allioli* in München — Hauptpastor Dr. *Alt* in Hamburg — Consistorialrat in Frankfurt a. M. — Professor Dr. *Aschbach* ebendas. — Rabbiner Dr. *Auerbach* in Generalsuperintendent und Consistorialrath *Bank*, Abt von Michaelstein, Ritter d braunsch. Ordens Heinrichs des Löwen, in Wolfenbüttel — Dr. *Johann Wilhelm* von Limburg — Professor Dr. *Jos. Beck* in Freiburg im Breisgau — Consistorialrat Dr. *Beckhaus* in Marburg — Professor Dr. *Bercht* in Frankfurt a. M. — Dr. *Bernhard* der Landesbibliothek in Cassel — Oberappellationsgerichtsrath Dr. *Bickell*, Ritter hess. Ordens vom goldenen Löwen — Herr von *Biedenfeld* in Weimar — Simon B des Régénérateur in Straßburg — Oberpräsident Graf von *Blücher-Altona*, Großk Dannebrog-Ordens, Präsident des Consistoriums in Altona — Generalsuperinter me Kirchenrath Dr. *Böckel*, großherzogl. Oberhofprediger in Oldenburg — Domk geistl. Rath Dr. *Bohn* in Frankfurt a. M. — Domkapitular *Brand* in Hildesheim — tendent *Breiger* in Harburg — Domkapitular und geistl. Rath Professor Dr. *Brenn* Christian *Brentano-Laroche* in Frankfurt a. M. — Domkapitular und geistl. Rath B Klerikalseminars in Limburg — C. M. *Brug*, Ritter des päpstl. Ordens vom golden teur der Sion, in Augsburg — Dr. *Bruschius* in Dresden — Dr. *Büdingen* in Cassel *Bülow*, Erb-Landmarschall des Herzogthums Sachsen-Lauenburg, kön. dän. Kam des Dannebrog-Ordens, ritterschaftl. Assessor bei dem Consistorium in Ratzebu assessor *Burchard* in Coblenz — Professor Dr. *Credner* in Gießen — Dr. *Creizer*

BeataMache 4 months ago

Allioli

Joseph Franz von Allioli, Theologieprofessor in München (GND <http://d-nb.info/gnd/119101564>) bestritt, der Veröffentlichung seines Namens in der Ankündigung zugestimmt zu haben. Mehr dazu: STEINSDORFER, HELMUT (1976): Eine deutschsprachige interkonfessionelle Kirchenzeitung und die Römische Kurie im Jahre 1837. <http://www.perspectivia.net/content/publikationen/qfiab/55-56-1976/0276-0314>

Reply Link

BeataMache 4 months ago

Staudenmaier


<http://d-nb.info/gnd/178196746>

Reply Link

nubuker 4 months ago

Franz Varrentrapp

Der Enkel des Verlagsgründers hieß ebenfalls Franz Varrentrapp (1815–1877). Er war Chemiker und Verleger.



Seine GND: <http://d-nb.info/gnd/117350559> (Bildnachweis)

Abb.: Hypothes.is: Annotation eines PDF

Ausblick

Dass das Konzept "Web-Annotation" weithin auf Interesse stößt, mag man auch daran ablesen, dass es unter Federführung des Projekts "Hypothes.is" gelungen ist, eine Kooperation zahlreicher namhafter Akteure, "world's key scholarly publishers, platforms, libraries, and technology organizations", ins Leben zu rufen.²⁴

Für die wissenschaftliche Nutzung stellt sich die Herausforderung, dass Web-Annotation nicht durch konkurrierende Plattformen und den nicht ganz einfach zu leistenden Spagat zwischen eher generischen und eindeutig fachspezifischen Use Cases auf Kosten der Interoperabilität in zahllose inkompatible Nischen zerfällt. Hier sind vor allem Standards wie das Open Annotation Data Model²⁵, maschinenlesbare Schnittstellen sowie die konsequente Unterstützung von kontrollierten Vokabularien und Normdaten gefragt, die die Einbettung in das "Semantic Web" erlauben.

²⁴ Annotating All Knowledge — A new open layer is being created over all knowledge <https://hypothes.is/annotating-all-knowledge/>

²⁵ <http://www.openannotation.org/>

Digitale Annotationen und das Semantic Web

Digitale Annotationen im Web bekommen mit der Entwicklung des Semantic Web eine tragfähige und nachhaltige technologische Basis. Seit der Entstehung des Social Web werden Internetnutzern Plattformen zur Verfügung gestellt, um Information im Web zu veröffentlichen und zugänglich zu machen, mit anderen zu kommunizieren und zu kollaborieren. Webnutzer haben die Möglichkeit, plattformübergreifend zu publizieren, Inhalte weiter zu verwenden und zu vermischen. Nicht zuletzt im Kontext kollaborativer semantischer Anwendungen gilt Standardisierung dabei als Grundlage für die Zugänglichkeit und Nutzbarkeit dieser Informationen. Das Semantic Web bietet einen Rahmen, der es möglich macht, Information zu formalisieren, auf Basis eines einheitlichen Vokabulars sowie gemeinsamer Konzepte und Modelle disziplin- und plattformübergreifend zu teilen und wiederzuverwenden. Mit dem Semantic Web rückt die Entwicklung intelligenter Kollektionen sowie die nutzergenerierte Komposition, Integration und Aggregation von Inhalten in den Vordergrund der Betrachtung.

Semantische Web-Technologien ermöglichen einen neuen Ansatz, Information zu verwalten und zu erschließen und ein Netz herzustellen, in welchem Objekte direkt aufeinander verweisen. Der Prozess umfasst dabei die Generierung und Nutzung von Metadaten auf zwei Ebenen, einerseits zur Beschreibung der Dokumenteninhalte, ferner der Entitäten innerhalb eines Dokuments (Davies et al, 2007). Insbesondere charakteristisch für das webbasierte semantische Annotieren ist es, diese Metadaten nicht nur mit eindeutigen Identifikatoren zu versehen, um Interoperabilität und eine eindeutige Referenzierbarkeit der Annotationen zu gewährleisten, sondern darüber hinaus Verknüpfungen zwischen Ressourcen herzustellen und diese zu typisieren. Als Grundlage greifen die fundamentalen Konzepte des Semantic Web: Jeder kann etwas zu einem Thema sagen (AAA Slogan), eine Wissensbasis ist immer potentiell unvollständig, wobei alles, was nicht modelliert wurde, als nicht existent angenommen wird (open world/closed world); darüber hinaus kann jede Entität mit mehr als nur einem Bezeichner beschrieben werden (non-unique naming) und es ist ein organisches Wachstum zu erwarten (network effect). In diesem Zusammenhang beeinflussen die spezifische Zielsetzung des annotierenden Nutzers sowie das Werkzeug selbst den Annotationsprozess. Schwierigkeiten entstehen insbesondere bei der Formalisierung von Wissen mit Hilfe von semantischen Editoren und browserbasierten Werkzeugen sowie bei der Darstellung und Nutzung semantischer Information für Anfragen und das Browsen.

Mit RDF/OWL wird ein standardisierter Formalismus als Datenmodell zu Grunde gelegt, der es erlaubt, Information über das Web zu verteilen und zu verknüpfen. Der Prozess des semantischen Annotierens von Web Ressourcen umfasst, wie bereits beschrieben, in der Regel das Hinzufügen von Information, ohne dabei die ursprüngliche Ressource zu verändern. Zudem gibt es drei verschiedene Ansätze, die entstandenen Annotationen zu speichern. Der einfachste Ansatz ist, diese zusammen mit der annotierten Ressource lokal vorzuhalten. Alternativ besteht die Möglichkeit, die Annotationen separat in einem Dokument zu speichern und zusammen mit der Ressource in einem Browser zu laden. Schließlich gibt es noch die Option, Annotationen direkt in eine Webseite einzubetten. Wichtige Werkzeuge und Entwicklungen in diesem Bereich sind:

Annotea ist als W3C Projekt der De-facto-Standard zur kollaborativen Annotation beliebiger Dokumente oder dem Teilen von Dokumenten. Die Annotationen werden parallel auf einem Annotations-Server gespeichert und zusammen mit dem Dokument geladen. Ein Zeiger verlinkt dann die Annotation und den entsprechenden Teil im Dokument.

Neonion unterstützt ebenfalls das kollaborative Annotieren basierend auf Annotator, einer JavaScript Bibliothek zur Integration und Speicherung authentifizierter Annotationen. Neonion ermöglicht die Annotation von Dokumenten auf unterschiedlichen Ebenen und integriert zudem automatische Annotationsdienste.

Catch integriert verschiedene Annotationswerkzeuge basierend auf Annotator und aggregiert die gespeicherten Annotationen angebundener Server.

Insbesondere die getrennte Speicherung der Annotationen hat diverse Vorteile und bietet mehr Flexibilität hinsichtlich des Zugriffs, der Nutzung, Verarbeitung und Auswertung dieser Annotationen. Die verschiedenen Ansätze erfüllen vorrangig spezifische Benutzeranforderungen und kommen zunächst ohne die Nutzung von Ontologien aus. Allerdings können Ontologien nicht nur genutzt werden, um Annotationen zu generieren, sondern auch um diese einer Validierung und Qualitätskontrolle zu unterziehen und Schlussfolgerungen über den Annotationsgegenstand zu ziehen (Hunter, 2009). Somit ist es auch möglich, fachspezifische Annotationen disziplinübergreifend miteinander zu verlinken. Es gibt zahlreiche Beispiele für das ontologiebasierte Annotieren (Handschuh und Staab, 2003). Eine aktuell vielversprechenden Lösung ist wie bereits erwähnt Pundit. Pundit ermöglicht Nutzern das Annotieren einer Webseite durch Hinzufügen strukturierter semantischer Auszeichnungen in RDF. Das Framework basiert auf dem Open Annotation Model (OAM), welches die konzeptionelle Basis für eine Verlinkung von Ressourcen liefert. Neben der Kollaboration beim Annotationsprozess ist es zudem mög-

lich, Annotationen anderer Nutzer weiterzuverwenden. Ein gemeinsames und erweiterbares konzeptionelles Modell zur Repräsentation der Annotationen ist somit grundlegend.

Während sich zunehmend erste Ansätze und Werkzeuge etablieren, ist weitere Forschungsarbeit als Basis für die Entwicklung kollaborativer Werkzeuge zur Unterstützung interdisziplinärer Forschungsarbeit sinnvoll und notwendig.

Text

Die schriftsprachliche Annotation schriftlicher (auch verschriftlichter) Texte kann die Vergabe von Metadaten, die textstrukturelle Annotation, Auszeichnungen auf der lexikalischen bzw. grammatischen Ebene sowie semantische Annotationen unterschiedlicher Tiefe bzw. Komplexität umfassen – je nach Gegenständen und Projektzielen bzw. Forschungsfragen. Sowohl die verschiedenen Gegenstände, die annotiert werden, als auch die unterschiedlichen Formen und Funktionen von Annotationen bilden ein inhaltlich und medial sehr heterogenes Spektrum. Auch im eingegrenzten Bereich von Schrifttexten (sowohl als Annotationsgegenstände als auch als Annotationsinhalte) ist die Vielfalt groß.

Differenzierungen verschiedener Textsorten bzw. –typen und –klassen (in der Linguistik) sowie Gattungen und Genres (in der Literaturwissenschaft) erlauben einen Überblick über den Facettenreichtum des Bereichs schriftsprachlicher Annotationsgegenstände. Unterschiedliche Kategorisierungen zeigen zugleich, dass es Überschneidungen zwischen den Formen gibt und trennscharfe Abgrenzungen oft schwierig sind. Textsorten-Kategorisierungen werden in der Linguistik anhand verschiedener Faktoren bestimmt. Beispielsweise wurde versucht, eine systematische Differenzierung durch Merkmalsbündel zu erreichen (vgl. z.B. Sandig 1975), wobei die Bestimmung bzw. Abgrenzung und Gewichtung von Merkmalen jedoch problematisch ist, oder durch pragmatische Unterscheidungen nach funktionalen oder situativen Gesichtspunkten (vgl. Heinemann 2000, Adamzik 2008), die jedoch auch keine eindeutige Systematisierung zulassen. Das Gefüge literarischer Gattungen – auf die Unterteilung nach Aristoteles Poetik zurückgehend, die auch von Goethe aufgegriffen wurde – wird ebenfalls als im Wandel begriffen. Auch bzgl. literarischer Texte wird auf unterschiedlichen Ebenen und nach verschiedenen Konzepten differenziert, bspw. grenzt man auch Genres voneinan-

der ab (vgl. z.B. Lamping 2009). Das Spektrum möglicher schriftsprachlicher Annotationsgegenstände umfasst also eine Formen-Vielfalt – von verschiedenen Sorten von Gebrauchstexten bis zu den unterschiedlichen Arten von Dichtung.

An Versuchen zur automatischen Erkennung von Textsorten und Gattungen wird geforscht (vgl. z.B. Mehler/Lobin 2004, Schöch 2014), aber in der TEI (Text Encoding Initiative) wurde die beschriebene Vielfalt im Hinblick auf die strukturelle Textauszeichnung bisher nur ansatzweise aufgearbeitet. Die TEI bietet spezielle Auszeichnungen für Versdichtung („Verse“), Verschriftlichungen von Theater und anderen Aufführungen („performance text“), Transkription gesprochener Sprache, Wörterbücher, Beschreibungen von Handschriften und Faksimiles, für den textkritischen Apparat in Editionen sowie für nicht-schriftliche Komponenten (vgl. Sperberg-McQueen / Burnard 2015). Standardisierte Textsorten- und Gattungsauszeichnungen sind (noch) nicht umgesetzt, ihre Standardisierung wäre aufgrund der heterogenen Systematisierungsansätze und Merkmalsbündel auch problematisch und nicht auf der rein strukturellen Ebene umsetzbar. Allerdings gibt es Arbeitsgruppen und Initiativen, die sich mit der Standardisierung bestimmter Textsortenauszeichnungen befassen, z.B. Briefe und Tagebücher (vgl. Stadler / Veit 2009). Für bestimmte Inhaltsebenen gibt es hingegen TEI-Standards – etwa für Namen, Orte, Zeiten. Und auch für Annotationen sind Standards vorgesehen – sowohl für linguistische Merkmale als auch für Annotationen im Sinne von Notizen, also auch semantische Annotations-Möglichkeiten jenseits der Textstruktur (vgl. ebd.).

Schriftsprachliche Annotationen von Schrifttexten können also lexikalische und grammatische Merkmale auszeichnen (Lemmatisierung, Tokenisierung, Part-of-Speech-Tagging usw.), aber auch für komplexere semantische Merkmale, Notizen und Kommentierungen genutzt werden. Sie können der individuellen Textarbeit dienen, aber auch öffentlich sein. Sie können die Basis automatisierter Verarbeitungsprozesse bilden (bspw. in Form von Markierungen, Kennzeichnungen lexikalischer oder grammatischer Einheiten), und in Form von Schemata bzw. Taxonomien Datenmodellierungen darstellen (vgl. McCarty 2004, S. 254-271). Annotationen können außerdem als Verlinkung zwischen digitalen Inhalten fungieren und somit intertextuelle Verweisungen und Kontextualisierungen explizit machen, aber auch selbst Ergebnisse wissenschaftlichen Arbeitens im Sinne von Mikropublikationen darstellen (vgl. Bradley 2012). Insofern besteht ein Zusammenhang zwischen Annotation, Textvernetzung und hypertextueller bzw. – medialer Produktion, also digitalem Schreiben bzw. Verfassen medial unterschiedlicher

Inhalte, aber auch ein Bezug zu digitalen Editionen, die annotierend kommentiert werden können (Annotation als Stellenkommentar und eigene (Para-)Textsorte) und an die angeknüpft werden kann.

Im Projekt ePoetics (Kooperation zwischen der Technischen Universität Darmstadt und der Universität Darmstadt) werden literaturtheoretische Schriften digitalisiert, annotiert und analysiert. Bei den Annotationsgegenständen handelt es sich um ‚Lehren der Dichtkunst‘, aber nicht im Sinne von normativen Regelpoetiken. Konkret umfasst das Corpus 20 Poetiken von 1770-1960, in denen literaturtheoretisches Wissen diskursiv verhandelt wird. Im Projekt steht der Begriff der Metapher im Mittelpunkt. Problemstellungen für die digitale Annotation mit dem Ziel der computergestützten Auswertbarkeit liegen bei solchen Texten auf mehreren Ebenen vor: Das jeweilige Metaphernverständnis muss differenziert erschlossen und die Komponenten der Begriffsbestimmung müssen trennscharf kategorisiert werden können. Eine eindeutige Zuordnung zu Ansätzen bzw. Denkschulen der antiken Rhetorik soll erzielt und Bezüge zu anderen Poetiken sollen erfasst werden. Beispiele aus der Primärliteratur sind eindeutig den jeweiligen theoretischen Aspekten zuzuordnen. Und schließlich müssen die Textebenen und Referenzstrukturen der Poetik explizit gemacht werden – also wo der Autor selbst theoretisiert, wo zitiert oder paraphrasiert wird, inwiefern dies kenntlich gemacht wird oder nicht und wo vom ursprünglichen Text abgewichen wird. Dies wird durch die Annotation nach einem komplexen Schema umgesetzt. Die Annotationen werden einerseits in TEI-konformen XML-Dateien publiziert, andererseits aber auch als Grundlage von computergestützten Analysen und Visualisierungen genutzt. Die Inhalte der Annotationen umfassen also Angaben zu Referenzrelationen, konkrete Namen und Werke zitierter Autoren sowie Komponenten sprach- und literaturtheoretischer Kategorien. Als Werkzeuge wurden das Wagsoft UAM Corpus Tool sowie das TextGrid-Laboratory als kollaborative XML-Annotationsumgebung genutzt. Publiziert wird das Corpus im TextGrid/DARIAH-Repository sowie im Deutschen Textarchiv der Berlin-Brandenburgischen Akademie der Wissenschaften.

Das Corpus des Projekts Vernetzte Korrespondenzen (Kooperation zwischen dem Trier Center for Digital Humanities, dem Deutschen Literaturarchiv Marbach und der Martin Luther-Universität Halle-Wittenberg) besteht aus vielen Briefen unterschiedlichster Urheber, die an eine Vielzahl von Adressaten gerichtet sind. Hinter den Briefwechseln des Projekts verbirgt sich also nicht nur die Korrespondenz einiger Weniger untereinander.

Das Exil bildet dabei das verbindende Element. Dies bringt spezielle Herausforderungen für das Annotieren mit sich. Einerseits müssen die Briefe TEI-konform ausgezeichnet werden, andererseits bildet das Anreichern mit Geokoordinaten, Normdaten zu Personen, ihren Berufen oder Konfessionen und bibliografischen Angaben einen Schwerpunkt des Projekts, was häufig nur mit großem Recherche-Aufwand zu realisieren ist. Als Tool wird FuD (<http://www.fud.uni-trier.de/de/genutzt>) genutzt. Darüber hinaus soll aus dem Projekt ein manuell erstellter Thesaurus zur automatischen Verschlagwortung einzelner Briefe hervorgehen sowie eine interaktive Plattform zu Thema und Corpus. Die Annotationen dienen der Erforschung und Visualisierung sozialer, räumlicher, zeitlicher und thematischer Netze in Briefkorpora.

Im Projekt Natur & Staat, Teil des Graduiertenkollegs Knowledge Discovery in Scientific Literature (KDSL, Technische Universität Darmstadt), wird ebenfalls zur Metapher geforscht – allerdings steht dort nicht der Metaphernbegriff selbst, sondern die Annotation von Metaphern in (populär-) wissenschaftlichen Texten Anfang des 20. Jahrhunderts im Mittelpunkt. Ausgangspunkt für die digitale Annotation von Metaphern in diesen Quellentexten sind Metapherntheorien – insbesondere nach Lakoff/Johnson, Max Black und Petra Gehring (Gehring / Gurevych 2014). Diese unterschiedlichen theoretischen Modelle bedingen Metaphernannotationen auf verschiedenen Ebenen, theoretische Herangehensweisen werden jeweils probeweise operationalisiert. Es werden Text-Abschnitte annotiert, die Metaphern-Teile darstellen, sowie die Relationen zwischen diesen Teilen (z.B. zwischen Source und Target, Bild-Spender und Bild-Empfänger). Als kollaboratives Annotationswerkzeug wird WebAnno (<https://www.ukp.tu-darmstadt.de/software/webanno/>) genutzt. Die Annotationen sollen vor allem Kontextbrüche erfassen, die als Indikatoren für Metaphorik verstanden werden. Sie sollen als Trainingsdaten für einen Classifier genutzt werden, der automatisch Metaphern erkennen soll (weitere Ausführungen im folgenden Abschnitt).

Projekte der Infrastruktur Webanno

Webanno ist eine initial für CLARIN entwickelte Infrastrukturlösung zur textuellen Annotation (Castilho 2014). Sie verfolgt einen ‚general purpose‘-Anspruch, der zugleich webbasierte, kollaborative Annotations-Arbeit eröffnet. Diese Infrastrukturlösung liegt – auch mittels Communityfeedback – im dritten major-release vor (Apache Software License / ASL, version 2), wird aktiv gepflegt, bietet Demo-Zugänge und ist sehr gut bei

github dokumentiert.²⁶ Bei Youtube stehen zudem Tutorials zur Verfügung, die in die Arbeitsweise einführen²⁷.

Webanno zielt vor allem auf linguistisch fundierte Anwendungsfälle mit einem bis zum zweiten Release insbesondere diesbezüglich hilfreichen Funktionsumfang, der sich auf vordefinierten Annotations-, Schichten‘ bewegt (morphologische, syntaktische, semantische sowie part of speech, lemmata, named entities, dependency relations, coreference chains) sowie benutzerdefiniert eröffnete zum Einsatz bringen lässt (Yimam et al 2014). Strukturell setzt Webanno ein differenzierbares Nutzer-Rollenmodell um (Basis: „Manager“, „Annotators“, „Curator“; Support für Inter-Annotator-Agreement‘), wobei neben (IAA-/progress-)Monitoring zusätzlich das Outsourcen segmentierbarer Aufgaben mittels einer Crowdsourcing-Schnittstelle unterstützt wird (crowdfunder, ggf. Amazon mechanical-turk) . Annotation erfolgt hauptsächlich mittels Markieren textueller spans, relationszwischenspannsowie chains, ab Webanno 3 sind zudem generische Annotationen (z.B. Focusausdruck und Kontext) möglich. Innovativer Aspekt der Lösung sind Optionen zur Automatisierung von Arbeitsschritten mittels nutzer- und taskspezifisch zuschaltbarer machine-learning-Komponenten, die nach jeweiligem Training z.B. Annotations-Vorschläge unterbreiten können (Yimam et al 2014).

Auf die Offenheit der Architektur und bestehende Schnittstellen (neben DKPro-orientierten Annotationsschichten und crowdsourcing u.a. TCF und TEI) bauten in der Vergangenheit z.B. Projekte zur Vorbereitung von Daten zur deutschen Named Entity Recognition, der semantischen Annotation eines Dänischen Referenz-Korpus und – jüngst das im letzten Abschnitt genannte Projekt ‚Natur und Staat‘ (KDSL)²⁸, das mittels Webanno ein 9-bändiges Korpus sozialdarwinistischer Texte aus der Zeit der Jahrhundertwende zum 20. Jh. auf Metaphernaufkommen annotiert (e-Metapher-Algorithmus: Gehring / Gurevych 2014), daneben wird der ‚-ismen‘-Gebrauch annotiert. Hierdurch werden in beiden Fällen erfolgreich Algorithmen zur automatischen Detektion dieser beiden sprachlichen Formen trainiert.

Insbesondere das KDSL-Projekt deutet in seiner WebAnno-Nutzung bereits in Richtungen und Perspektiven, die gegenwärtig weitgehende Desiderata digitaler Annotationsumgebungen identifizierbar werden lassen: Projektorientierte Inputs an Annotations-

²⁶ <https://webanno.github.io/webanno/> Home <https://webanno.github.io/webanno/releases/2.3.0/docs/user-guide.html> user-guide

²⁷ <https://www.youtube.com/channel/UCew3TzVAw4MHIt0zAHFlcxQ> webanno-tutorials

²⁸ <https://www.kdsl.tu-darmstadt.de/de/kdsl/research-program/natur-und-staat-eugenics-and-metaphor/> Home

Infrastrukturen, die zugunsten der breit ausdifferenzierten geisteswissenschaftlichen Forschungslandschaft paradigmatisch eine mittels Annotationen verfahrenende Strukturierung von Texten kategorial weniger generisch denn initial an strikt situierten Zugriffsweisen, Semantiken und Ontologien disziplinärer und methodologischer Felder ausrichten und dadurch von Beginn an aktiv (z.B. auch im Kontext von anschließenden tagging-Prozessen) jeweils disziplinär / methodisch singuläre Herangehensweisen und Verfahrenstypen digital umsetzbar und auch hierdurch nachhaltig bewahrbar gestalten.

Bild

In den digitalen Geisteswissenschaften finden die Guidelines der Text Encoding Initiative (TEI) vielfache Anwendung. Auch wenn die TEI sich primär die Aufgabe gestellt hat, einen "Standard zur Repräsentation von Texten in digitaler Form" (tei-c.org) zu entwickeln und zu etablieren, enthalten die Richtlinien dennoch einige Elemente, die sich – wenn auch in begrenztem Maße – für die Annotation von Bildern eignen. `figure`, `graphic` und `figDesc` ermöglichen Bilder in Texten zu lokalisieren und zu beschreiben. Das Instrumentarium der TEI ermöglicht also durchaus Bild und Text in Beziehung zu setzen, doch immer vom Text her gedacht. Eine eigenständige Repräsentation der visuellen Aspekte oder des Objektcharakters (jenseits der Textualität) kann (und will) die TEI nicht leisten. Seit einigen Jahren widmet sich die Special Interest Group (SIG) "Text and Graphics" verstärkt den graphischen Komponenten. Zweck dieser SIG "will be to discuss, develop, and document various strategies and best practices for implementing the digital facsimile structures available in TEI P5, especially as relates to graphically rich texts, and for developing methods to extend the TEI when existing structures are not suitable."²⁹ Zudem gibt es mehrere Initiativen, die versuchen, den spezifischen Herausforderungen, die "bildlastige" Genres (Comic, graphic novels) an einen Auszeichnungsstandard stellen, gerecht zu werden, wie etwa das GraphML Format³⁰ oder die an die TEI angelehnte Comic Book Markup Language³¹.

Die Digitale Infrastruktur TextGrid/DARIAH bietet mit dem Text Bild Link Editor ein Werkzeug an, mit dem beliebige Textsegmente mit Bildausschnitten verknüpft werden können. "Der Text-Bild-Link Editor (TBLE) im TextGridLab erlaubt es, Teile von Bildern mit Elementen eines strukturierten Texts zu verbinden. Im Bild können Regionen mit

²⁹ <http://www.tei-c.org/Activities/SIG/Graphics/>

³⁰ <http://graphml.graphdrawing.org>

³¹ <http://dcl.slis.indiana.edu/cbml/>

Rechtecken oder Polygonen markiert werden. Dann können im Text XML-Elemente oder beliebige Bereiche ausgewählt und mit der Bildregion verlinkt werden. Die entstandenen Links können dann als eigenständiges Objekt im TextGridRep gespeichert und wieder aufgerufen werden. Mithilfe des TBLE können beispielsweise Transkriptionen von Manuskripten Wort für Wort oder Zeile für Zeile mit den gescannten Bildern der Seiten verbunden werden. Es ist auch möglich, eigene Notizen mit Bildern zu verlinken und auf diesem Weg Bild-Annotationen zu erstellen." (Casties/Wintergrün 2015, S. 161f.)

Bei der OpenSource Software DigiLib³² stand von Anfang an die (kollaborative) Arbeit mit Bilddigitalisaten sowie grossen Bildermengen und die Anforderungen der bildbasierten Wissenschaften im Zentrum der Entwicklung. DigiLib zeichnet die Fähigkeit aus "Markierungen auf dem Bild zu setzen und in Form einer URL zu speichern. Diese Art der Annotation existiert nur lokal im Browser des Benutzers und enthält keine weiteren Informationen. Dennoch lässt sich eine solche URL zusammen mit einem Kommentar in einer E-Mail verschicken oder zur Illustration in einem PDF- oder HTML-Dokument verwenden. [...] Die Möglichkeit, mit der digilib-URL einen bestimmten Aspekt oder ein Bilddetail zu markieren und so zu annotieren, hat die Forscher von Anfang an begeistert. Sehr schnell kam in der Folge der Wunsch auf, mit der Annotation gleich einen Text zu verbinden, um sich beispielsweise in Gruppen über Bilder auszutauschen, ebenso wie der Wunsch, Annotationen anderer Benutzer zu sehen, die das gleiche Bild annotiert haben. [...] digilib bietet seit kurzem ein Annotations-Plug-in, mit dem Markierungen auf einem Bild mit kurzen Texten versehen und auf einem Annotationsserver gespeichert werden können." (Casties/Wintergrün 2015, S. 158f.)

Auch die Lüneburger Entwicklung Hyperimage ermöglicht die kollaborative Forschung direkt am Bild ohne den Umweg der textuellen Bildbeschreibung. Hyperimage ermöglicht sozusagen Fußnoten direkt an Bilder, Bildauschnitte zu knüpfen, sowie die Verknüpfung zwischen mehreren Bildern. "HyperImage realisiert die bildhafte Fußnote am Bild. Das bedeutet, dass präzise markierte Bilddetails mit anderen Bilddetails, Bildsammlungen und Texten hypertextuell verknüpft werden können. [...] Details auf Bildern werden bislang hingegen konventionell lokalisiert, indem Position und Identität entweder gestisch (Zeiger, Pfeile, Kreise) oder aber sprachlich (Vorder-, Mittel-, Hintergrund bzw. links/rechts, oben/unten) bezeichnet und abgegrenzt werden. [...] So reduzieren

³² <http://digilib.sourceforge.net>

sich Bildverweise typischerweise auch immer darauf, Bilder nur als Ganzes zu adressieren. Mit HyperImage können beliebig viele Details innerhalb eines Bildes präzise markiert und beschrieben sowie Annotationen des Corpus untereinander verlinkt und über Indizes erschlossen werden".³³ Zahlreiche Projekte wurden mit Hyperimage realisiert, etwa der Nachlaß der Künstlerin Anna Opermann³⁴, oder die mittelalterliche Ebstorfer Weltkarte.³⁵

Eine spezifische Anwendung von HyperImage wurde im BMBF geförderten eHumanities Projekt Relationen im Raum angewandt. Ziel des Projekts war die Analyse und Visualisierung räumlicher Relationen zwischen Grabmalen jüdischer Friedhöfe. Hierzu wurden Lagepläne von Friedhöfen, auf denen die einzelnen Grabmale verzeichnet sind, digitalisiert und mit umfangreichen Datenbanken der Judaistik und historischen Bauforschung hinterlegt. Über ein Suchinterface werden die Projektdatenbanken erschlossen und im Softwaresystem Hyperimage visualisiert werden.³⁶

Semantisch-topologische Annotationen von Bilddigitalisaten ermöglicht das im Rahmen des DARIAH-DE Cluster 6.3 realisierte Werkzeug Sem2notes: "SemToNotes möchte im Besonderen zeigen, dass sich Bildinhalte nicht nur Mithilfe von textuellen Annotationen erklären und analysieren lassen, sondern auch durch graphische (nicht-textuelle) Annotationen".³⁷ Ausführlich wird das Werkzeug sowie die Kooperationen im Bereich "Digitale Schriftkunde" und "eCodiology" sowie die Einbettung in bestehende Workflows, wie die TextGridLab Umgebung und den Heidelberger Digitalisierungsworkflow im DARIAH-DE M.6.3.1.

Video

Ebenso wie digitale Bilder bzw. digitale Repräsentationen von Gemälden, Objekten etc. annotiert werden können, können es auch die einzelnen Bilder – genannt Frames – eines digital vorliegenden Videos. Der wesentliche Unterschied zwischen einem Bild und einem so genannten Bewegtbild besteht in der zeitlichen Dimension, d.h. darin, dass es sich bei letzterem um eine Abfolge von Bildern handelt. Im Gegensatz zur

³³ <http://www.leuphana.de/institute/icam/forschung-projekte/hyperimage.html>

³⁴ http://www.uni-lueneburg.de/hyperimage/HI_Kunsthalle/

³⁵ <http://www2.leuphana.de/ebskart/>

Weitere Projekte: <http://www.leuphana.de/institute/icam/forschung-projekte/hyperimage.html> und <http://hyperimage.ws/de/>.

³⁶ <https://wiki.de.dariah.eu/display/RIRPUB/RiR>

³⁷ DARIAH-DE M 6.3.1 – Integration, 2016, S. 7.

Filmtechnik, bei der die einzelnen Bilder – fixiert auf einem Celluloid-Streifen – nacheinander abgespielt werden, um einen Bewegungsablauf zu suggerieren, baut sich bei der Videotechnik allerdings jeder Frame halbzeilenweise auf, so dass ein sanfter Übergang zwischen den Videobildern geschaffen wird. Aus diesem Grund ist es etwa möglich ein Video in Zeitlupe abzuspielen, ohne dass ein signifikanter Bruch erkennbar wäre. Es verwundert folglich nicht, dass Hyper Slow Motion ein probates Mittel vieler Videokünstler darstellt (vgl. hierzu weiter Rebentisch 2003, 195).

Die zeitliche Dimension eines Videos bedeutet in der Konsequenz für die Annotationspraxis, dass nicht nur die einzelnen Frames eine Annotation erfahren können, z.B. in Form von Schlagworten, Transkriptionen bzw. Untertiteln des Gesagten, Beschreibungen, Georeferenzierungen oder einer Verlinkung dargestellter Inhalte auf ikonographisch verwandte Bildmotive aus Hoch- und Populärkultur, sondern dass auch ganze Abschnitte eines Videos, z.B. bestimmte Einstellungen, annotiert werden können – und zwar wiederum mit ganzen Abschnitten, nicht nur mit Stills. Eine Besonderheit stellt es hierbei dar, wenn einzelne Abschnitte eines Videos wechselseitig aufeinander verweisen, so dass neben externen Bezügen auch selbstreferentielle Praktiken aufgedeckt werden können, denen im postmodernen Diskurs eine bedeutende Rolle zukommt.

Voraussetzung für die Annotation von Bewegtbildern ist ein entsprechendes Datenbanksystem wie es etwa Pan.do/ra bereitstellt. Matthias Arnold, Eric Decker und Thorsten Wübbena zeigen anhand einer Analyse des von Tarsem Singh umgesetzten Musikvideos zu R.E.M.'s *Losing My Religion* (1991) auf, wie die digitale Annotationstechnik eine bild- bzw. kunstwissenschaftliche Medienanalyse unterstützen kann, sich Informationstechnologie und Geisteswissenschaften also im Sinne der Digital Humanities fruchtbar verknüpfen und so einen Mehrwert erzielen, der über eine bloße Transformation analoger Methodiken ins Digitale hinausgeht (vgl. Arnold, Decker und Wübbena 2015). Um die vielfältigen kunst- und kulturgeschichtlichen Bezüge des beispielhaft ausgewählten Clips herausarbeiten zu können, muss mithilfe einer entsprechenden Annotationsplattform nicht länger der Umweg über Notizen in Textform gegangen werden, sondern es kann unmittelbar mit dem audiovisuellen Material selbst gearbeitet werden. Ein Vorteil des webbasierten Ansatzes von Pan.do/ra besteht hierbei in der zeit- und ortsabhängigen Verfügbarkeit der zu bearbeitenden Inhalte, so dass mehrere Personen übers Netz zusammenarbeiten können; ein anderer besteht darin, dass die Ergebnisse intuitiv nachnutzbar sind, insofern die Software ein "exploratives Sehen" (ebd.) ermög-

licht, da Benutzer "an den entscheidenden Punkten auf die Bezüge hingewiesen werden, nach denen sie nicht recherchieren konnten, weil sie schlicht nichts davon wussten." (ebd.) Aufgrund des komplexen Rechtemanagements der Software ist es hierbei möglich, die Annotationen nach Autoren zu filtern und den Annotationsprozess nachzuvollziehen. Darüber hinaus besteht die Möglichkeit Ausschnitte zu exportieren, um sie etwa in digitale Publikationsformate einzubetten.

Neben Funktionen zur manuellen Annotation von Videoclips, ist eine automatisierte Repräsentation des gesamten Videos in der digitalen Softwareumgebung von Relevanz, um sich einen Überblick über das gesamte Material zu verschaffen und hierin zu browsen. Ein bislang unerwähntes, aber nichtsdestoweniger relevantes Feature von Pan.do/ra stellt daher die Timeline dar, die in den vier unterschiedlichen Modi anti-alias, slit-scan, keyframes und waveform angeboten wird. Um die visuelle Struktur eines Videos zu erfassen, sind insbesondere die Ansichtsmodi anti-alias und slit-scan von großem Interesse, insofern hier die durchschnittlichen Farbwerte eines Frames dargestellt bzw. ... und so Schnitte sichtbar gemacht werden. Dergestalt ist die Timeline ein Mittel, das Marc Davis in seiner Untersuchung eines frühen Videoarchiv-Prototyps als Videogramm bezeichnet und das als visuelle Signatur eines Videos verstanden werden kann, da es trotz des temporalen Charakters eines Videos dessen Struktur auf einen Blick wiedergibt: "A videogram is made by grabbing a center strip from every video frame and concatenating them together. [...] the videogram represents the fine temporal resolution of the dynamics of the video with a reduced spatial resolution." (vgl. Davis 1995, 8)

Helligkeit, Sättigung etc. einzelner Frames können analysiert und Schnitte erkannt werden. Es können folglich Durchschnittswerte für das gesamte Video berechnet und als Metadaten bereitgestellt werden, diese können aber auch, wie eben geschildert, in Form einer Timeline visualisiert werden. Derartige automatisierte Analysen und Visualisierungen des Videomaterials sowie Weiterentwicklungen im Bereich des "(semi-)automatische[n] Erkennen[s] und Verfolgen[s] von Gesichtern, Gegenständen bzw. Mustern im Film" (vgl. nochmals Arnold, Decker und Wübbena 2015) sind wünschenswert und denkbar; ebenso ist es die weitere Einbindung kontrollierter Vokabulare, um den manuellen Annotationsprozess und damit die kollaborativ-wissenschaftliche und nachnutzbare Auseinandersetzung mit Bewegtbildern und der ihnen eigenen Sprache zu unterstützen.

Objekte

In der Archäologie gibt es zahlreiche Formate, Methoden und Verfahren der Objektannotationen. Diese sind im DARIAH-DE I Report unter dem Titel "Fachspezifische Empfehlungen für Daten und Metadaten" ausführlich vorgestellt worden.³⁸

Noten

Auf digitale Annotationen in der Musikwissenschaft wird am Beispiel des MEI Score Editor im DARIAH-DE Meilenstein 6.3.1³⁹ eingegangen.

³⁸ <https://wiki.de.dariah.eu/pages/viewpage.action?pageId=20058856> Wiki. Die Veröffentlichung dieses Reports in den DARIAH Working-Papers steht an.

³⁹ <> DARIAH-DE M 6.3.1 – Integration, 2016, S. 5f.

Literatur

Allgemein

Borek, Luise; Dombrowski, Quinn; Perkins, Jody; Schöch, Christoph (2015): TaDiRAH – Taxonomy of Digital Research Activities in the Humanities. Release Version 0.5.3, <https://github.com/dhtaxonomy/TaDiRAH>.

Webannotation

Harald Lordick (2014): »Lazarus Bendauid ?annotiert?«, in: Deutsch-jüdische Geschichte digital, 21. September 2014, <http://digid.hypotheses.org/360>.

Harald Lordick (2015): »Notizen zum Blog als Web-Annotation«, in: Deutsch-jüdische Geschichte digital, 30. Januar 2015, <http://digid.hypotheses.org/532>.

Semantic Web

Davies, J.; Struder, R. und Warren, P. (2007) "Semantic Web Technologie: Trends and Research in ontology-based Systems". San Francisco: Wiley.

Handschuh, S. und Staab, S. (2003) "Annotation for the Semantic Web". Amsterdam: IOS Press.

Hunter, J. (2009) "Collaborative semantic tagging and annotation systems". In Journal of Annual Review of Information Science and Technology, Vol. 43, No. 1, S. 1-84.

Text

Adamzik, Kirsten (2008): Textsorten und ihre Beschreibung. In: Nina Janich (Hg.): Textlinguistik – 15 Einführungen. Tübingen. S. 145-176.

Bender, Michael; Alscher, Stefan: Auf der Suche nach dem goldenen Baum. Digitale Annotation als Erkenntnisprozess und "tertium comparationis" – am Beispiel der Auszeichnung des Metaphernbegriffs in Poetiken. (akzeptiert, erscheint vorauss. Mitte 2016 in der Zeitschrift für digitale Geisteswissenschaften ZfDG)

Bradley, John. 2012. Towards a Richer Sense of Digital Annotation: Moving Beyond a „Media“ Orientation of the Annotation of Digital Objects. Digital Humanities Quarterly 6 (2). <http://www.digitalhumanities.org/dhq/vol/6/2/000121/000121.html>. Zugegriffen: 31. Dezember 2015.

Castilho, Richard Eckart de; Biemann, Chris; Gurevych, Iryna; Yimam, Seid Muhle 2014: "WebAnno: a flexible, web-based annotation tool for CLARIN" In: Proceedings of the CLARIN Annual Conference (CAC)

CLARIN ERIC, October 2014. <https://www.clarin.eu/content/papers-posters-and-demos-cac2014>. Zugegriffen: 16.02.2015.

Gehring, Petra, Gurevych, Iryna (2014): Suchen als Methode. Zu einigen Problemen digitaler Metapherndetektion. In: Journal Phänomenologie 41/2014. S. 99-109.

Heinemann, Wolfgang (2000): Textsorte – Textmuster – Texttyp. In: Klaus Brinker u. a. (Hg.): Text- und Gesprächslinguistik. Ein internationales Handbuch zeitgenössischer Forschung. Bd.: 1, Berlin

Hildenbrandt, Vera; Kamzelak, Roland; Molitor, Paul; Ritter, Jörg (2015): „im Zentrum eines Netzes [...] geistiger Fäden“. Erschließung und Erforschung thematischer Zusammenhänge in heterogenen Briefkorpora. In: Datenbank-Spektrum 15, 1 (2015), S. 49–55.

Lamping, Dieter (Hg.) (2009): Handbuch der literarischen Gattungen. Stuttgart.

McCarty, Willard. 2004. Modeling – A Study in Words and Meanings. In A Companion to Digital Humanities, Hrsg. Susan Schreibman, Ray Siemens und John Unsworth, 254–271. Oxford: Blackwell.

Mehler, Alexander / Lobin, Henning (2004): Automatische Textanalyse

Systeme und Methoden zur Annotation und Analyse natürlichsprachiger Texte. Wiesbaden.

Sandig, Barbara (1975): Zur Differenzierung gebrauchsspezifischer Textsorten im Deutschen. In: Elisabeth Gülich und Wolfgang Raible (Hg.): Textsorten. Differenzierungskriterien aus linguistischer Sicht. 2. Aufl., Wiesbaden.

Schöch, Christoph (2014): Computergestützte literarische Gattungsstilistik. <http://www.germanistik.uni-wuerzburg.de/lehrstuehle/computerphilologie/mitarbeiter/schoech/cslg/>

Sperberg-McQueen, Michael / Burnard, Lou et al. (2015): TEI P5 Guidelines for Electronic Text Encoding and Interchange, <http://www.tei-c.org/release/doc/tei-p5-doc/en/Guidelines.pdf>

Stadler, Peter / Veit, Joachim (Hg.) (2009): Digitale Edition zwischen Experiment und Standardisierung. Musik – Text – Codierung. Beihefte zu Editio. Tübingen.

Yimam, Seid Muhle; Eckart de Castilho, Richard Eckart; Gurevych, Iryna; Biemann, Chris: "Automatic Annotation Suggestions and Custom Annotation Layers in WebAnno". Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pages 91–96, Baltimore, Maryland USA, June 23-24, <http://www.aclweb.org/anthology/P/P14/P14-5016.pdf>

Bild

Casties, Robert; Wintergrün, Dirk (2015). Bilder als Quelle in TextGrid. in: Neuroth, Heike; Rapp, Andrea; Söring, Sibylle (Hrsg.): TextGrid: Von der Community für die Community – Eine Virtuelle Forschungsumgebung für die Geisteswissenschaften. S. 153–163.

Video

Arnold, Matthias, Eric Decker und Thorsten Wübbena. 2015. "Losing My Religion" – Einsatz der Videoannotationsdatenbank Pan.do/ra in der kunstgeschichtlichen Analyse von Musikvideos, in: Grenzen und Möglichkeiten der Digital Humanities, hg. von Constanze Baum und Thomas Stäcker (= Sonderband der Zeitschrift für digitale Geisteswissenschaften, Bd. 1), URL: http://www.zfdg.de/sb001_018 (zuletzt aufgerufen am 25.01.2015)

Davis, Marc. 1995. "Media Streams: An Iconic Visual Language for Video Representation", in: Readings in Human Computer Interaction: Toward the Year 2000, hg. von Ronald M. Baecker, Jonathan Grudin, William A. S. Buxton und Saul Greenberg, San

Francisco: Morgan Kaufmann Publishers, Inc., S. 854-866, URL: http://coral.ufsm.br/ti-elletcab/Nusi/HiperV/Biblio/PDF/Davis_MediaStreams.pdf (S. 1-16) (zuletzt aufgerufen am 11.12.2015)

Rebentisch, Juliane. 2003. Ästhetik der Installation, Frankfurt am Main: Suhrkamp.